BERT:

BERT is an encoder-only transformer that processes input sequences bidirectionally, considering context from both left and right simultaneously It is pre-trained using two tasks: Masked Language Modeling (MLM), where random tokens are masked and predicted, and Next Sentence Prediction (NSP), which determines the relationship between two sentences BERT is designed for fine-tuning on downstream tasks like classification, question answering, and named entity recognition Variants like RoBERTa and DistilBERT improve performance and efficiency.

GPT :

GPT is a decoder-only transformer optimized for autoregressive tasks, predicting the next word based on previous words It is trained unidirectionally (left-to-right) using large-scale datasets. GPT excels in text generation, including dialogue, story writing, and code completion Versions like GPT-2 and GPT-3 scale up the model size and capabilities significantly It is commonly used for creative tasks and conversational AI.

Vision Transformer :

ViT adapts the transformer architecture for image classification tasks by dividing an image into patches, treating these patches as a sequence of tokens Each patch is embedded and processed through transformer layers to capture global relationships It demonstrated that transformers could perform comparably to convolutional neural networks (CNNs) with sufficient data and training ViT has inspired advancements in computer vision tasks like object detection and segmentation.

T5 Transformer :

T5 is an encoder-decoder transformer that reformulates all NLP tasks as text-to-text problems. For example, translation is treated as converting a source sentence into a target language, and summarization generates a summary from input text It is pre-trained on a diverse set of tasks to handle various use cases effectively T5's flexibility and unified format make it highly versatile for natural language understanding and generation tasks.

BART:

BART combines the bidirectional encoding of BERT with the autoregressive decoding of GPT It is pre-trained using a denoising autoencoder objective, where corrupted input sequences are reconstructed into original sequences This makes it robust for tasks like summarization, text generation, and machine translation BART is highly adaptable for fine-tuning on many sequence-to-sequence applications.

CLIP:

CLIP is a multimodal transformer designed to understand the relationship between images and textual descriptions. It learns by matching image-text pairs through contrastive learning, ensuring semantically similar inputs align in embedding space. CLIP has been used for tasks like zero-shot classification, image retrieval, and multimodal understanding, showing strong performance without task-specific fine-tuning.

Longformer:

Longformer is a sparse transformer designed for long-document processing, reducing the quadratic complexity of standard transformers. It introduces a mix of local and global attention mechanisms to efficiently capture dependencies across long sequences. Longformer is ideal for tasks like document summarization, question answering, and text classification over long inputs.

Informer:

Informer is a transformer tailored for time-series forecasting, focusing on long-term dependencies in sequential data. It uses a sparse self-attention mechanism to handle long sequences efficiently and reduces memory usage. Informer excels in applications like energy consumption prediction, anomaly detection, and weather forecasting.

DALL-E:

DALL-E is a generative transformer designed to create images from textual descriptions. By learning to associate textual input with visual representations, it can generate diverse and creative images based on input prompts. DALL-E has applications in art, design, and content creation, showcasing the potential of multimodal transformers in creative tasks.

Graphormer:

Graphormer adapts the transformer architecture for graph-structured data, such as molecular graphs or social networks. It incorporates structural information into the self-attention mechanism to process relationships between nodes effectively. Graphormer is used for tasks like predicting molecular properties, graph classification, and knowledge graph representation.